

Stochastic Ordering of Fading Channels Through the Shannon Transform

Adithya Rajan, Cihan Tepedelenlioğlu, *Member, IEEE*

Abstract

A new stochastic order between two fading distributions is introduced. A fading channel dominates another in the ergodic capacity ordering sense, if the Shannon transform of the first is greater than that of the second at all values of average signal to noise ratio. It is shown that some parametric fading models such as the Nakagami- m , Rician, and Hoyt are distributions that are monotonic in their line of sight parameters with respect to the ergodic capacity order. Some operations under which the ergodic capacity order is preserved are also discussed. Through these properties of the ergodic capacity order, it is possible to compare under two different fading scenarios, the ergodic capacity of a composite system involving multiple fading links with coding/decoding capabilities only at the transmitter/receiver. Such comparisons can be made even in cases when a closed form expression for the ergodic capacity of the composite system is not analytically tractable. Applications to multiple access channels, and extensions to multiple-input multiple-output (MIMO) systems are also discussed.

Index Terms- Ergodic capacity, fading, stochastic order, Shannon transform.

I. INTRODUCTION

Consider a flat fading channel with additive white Gaussian noise (AWGN), where the receiver has perfect channel state information (CSI). The maximum achievable rate of this system, when coding is applied across multiple independent channel realizations is known as the ergodic capacity, and is given by $\mathbb{E}[\log(1 + \rho X)]$, where $\rho \geq 0$ represents the average signal to noise power ratio (SNR) of the system, and ρX represents the instantaneous SNR random variable (RV). This expectation is also known as the Shannon transform of X [2, pp. 44], [3].

The authors are with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA. (Email: {arajan2, cihan}@asu.edu). This work was supported in part by the National Science Foundation under Grant CCF 1117041. Part of this work has appeared in [1].

In this work, a stochastic order which can be used to compare fading channels based on the Shannon transform of the instantaneous SNR is discussed. A fading channel is said to be better than another in the ergodic capacity order, if its corresponding ergodic capacity is bigger for all ρ . The proposed order is a kind of stochastic order on positive RVs. Stochastic orders in general find applications in economics [4], reliability analysis [5], and actuarial sciences [6]. A comprehensive exposition of stochastic orders can be found in [7]. Previously, the stochastic Laplace transform (LT) order, which compares the real-valued Laplace transforms of RVs has been used to compare two fading distributions and applied to comparing the average error rate of M -QAM modulations [8]. This can be explained by the fact that error rates of some modulations are non-negative integral mixtures of decaying exponentials, which can also be viewed as the Laplace transform. It has been shown in [8] that Laplace transform ordering of instantaneous SNRs implies ordering of ergodic capacities, but not conversely.

The ergodic capacity order presented in Section III of this paper is new to both stochastic ordering literature as well as information theory literature. Although this stochastic order was first introduced in [1], the current paper offers a detailed discussion of its properties, examples and extensions relevant to wireless communications, including the MIMO case. Further, some of the convergence properties of the Shannon transform are also studied. In this paper, many parametric fading distribution families such as the Nakagami- m , Rician and Hoyt are observed to have the property that the ergodic capacity is monotone with respect to the line of sight (LoS) parameter for each of these distributions. Consequently, the instantaneous SNR of these fading channels serve as examples of ergodic capacity ordered random variables. The properties of this stochastic order are useful in obtaining comparisons of the performance of systems involving multiple SNR RVs, as described in Section IV. For example, let $\{X_i\}_{i=1}^M$ and $\{Y_i\}_{i=1}^M$ be two sets of fading channels such that the ergodic capacity of X_i is less than that of Y_i , $i = 1, \dots, M$ at all SNR. Then, the properties of the ergodic capacity order provide the conditions under which a composite system consisting of $\{X_i\}_{i=1}^M$ as the component fading channels has a smaller ergodic capacity than that of a system with components $\{Y_i\}_{i=1}^M$. Such comparisons of ergodic capacities can be made even in cases when a closed-form expression is not available. A MIMO extension of the definition of the ergodic capacity order, which can be used to order positive semidefinite symmetric random matrices is given in Section V.

A. Notations and Conventions

The set of real numbers, positive integers and complex positive semidefinite symmetric matrices of size $n \times n$ are denoted by \mathbb{R} , \mathbb{N} , and \mathbb{S}_+^n respectively, while all other sets are denoted using script font. For a finite set \mathcal{B} the cardinality is denoted by $\text{card } \mathcal{B}$, while the indicator function is defined as $I(x \in \mathcal{K}) = 1$, if $x \in \mathcal{K}$ and 0, otherwise. For any measure $\mu(\cdot)$, $\mu(u)$ is used to represent $\mu([0, u])$. Vectors and matrices are denoted by boldface lower-case and upper-case letters respectively. For both the cases, $\|\cdot\|$ denotes the L^2 norm. The trace and determinant of a matrix \mathbf{M} are denoted by $\text{tr } \mathbf{M}$ and $\det(\mathbf{M})$ respectively. The identity matrix is denoted by \mathbf{I} . If $a_i \in \mathbb{R}$, $i = 1, \dots, N$, then $\text{diag}(a_1, \dots, a_N)$ is the diagonal matrix whose (i, i) element is a_i , $i = 1, \dots, N$. The i^{th} smallest eigenvalue of $\mathbf{A} \in \mathbb{R}^{N \times N}$ is denoted by $\lambda_i(\mathbf{A})$, $i = 1, \dots, N$. For a random variable X , $F_X(x)$ and $f_X(x)$ denote the cumulative distribution function (CDF) and the probability density function (PDF) respectively. $\mathbb{E}[g(X)]$ is used to denote the expectation of the function $g(\cdot)$ over the PDF of X . All logarithms are natural logarithms. We write $f_1(x) = O(f_2(x))$, $x \rightarrow a$ to indicate that $\limsup_{x \rightarrow a} (f_1(x)/f_2(x)) < \infty$.

II. MATHEMATICAL PRELIMINARIES

A. Completely Monotone Functions

A function $g : (0, \infty) \rightarrow \mathbb{R}$ is completely monotone (c.m.), if and only if it has derivatives of all orders which satisfy

$$(-1)^n \frac{d^n}{dx^n} g(x) \geq 0, \quad (1)$$

for all $x \geq 0$ and $n \in \mathbb{N} \cup \{0\}$, where the derivative of order $n = 0$ is defined as $g(x)$ itself. The celebrated Bernstein's theorem [9] asserts that, $g : (0, \infty) \rightarrow \mathbb{R}$ is c.m. if and only if it can be written as a mixture of decaying exponentials:

$$g(x) = \int_{[0, \infty)} \exp(-ux) \mu(du), \quad (2)$$

which is a Lebesgue integral with respect to a positive measure μ on $[0, \infty)$. It is straightforward to verify that c.m. functions are positive, decreasing and convex, and that positive linear combinations of c.m. functions are also c.m [9].

B. Stieltjes Functions

Stieltjes functions are a subclass of completely monotone functions, denoted by \mathcal{S} . A function $g : (0, \infty) \rightarrow [0, \infty)$ belongs to \mathcal{S} if and only if it can be written of the form

$$g(x) = a/x + b + \int_{(0, \infty)} (x + u)^{-1} \mu(\mathrm{d}u) , \quad (3)$$

where $a, b \geq 0$, and μ is a non-negative measure on $(0, \infty)$ which satisfies the convergence condition $\int_{(0, \infty)} (1 + u)^{-1} \mu(\mathrm{d}u) < \infty$. It is easy to show that any Stieltjes function is also a double Laplace transform of a non-negative function. A necessary and sufficient condition for $x \mapsto g(x) \in \mathcal{S}$ is that $x \mapsto (g(x^{-1}))^{-1}$ also belongs to \mathcal{S} [9, p. 66].

C. Bernstein Functions

A function $g : (0, \infty) \rightarrow \mathbb{R}$ is a Bernstein function, if and only if $g(x) \geq 0, \forall x > 0$, and $\mathrm{d}g(x)/\mathrm{d}x$ is c.m.. Equivalently, $g(x)$ admits the representation [9, p. 15]

$$g(x) = a + bx + \int_{(0, \infty)} (1 - \exp(-ux)) \mu(\mathrm{d}u) , \quad (4)$$

for some $a, b \geq 0$, where μ is a non-negative measure on $(0, \infty)$ satisfying $\int_{(0, 1)} \mu(\mathrm{d}u) + \int_{[1, \infty)} u \mu(\mathrm{d}u) < \infty$. The set of all Bernstein functions is denoted by \mathcal{BF} .

An important property is that the set \mathcal{BF} is closed under positive linear combinations: if $g_i \in \mathcal{BF}$, and $a_i \geq 0, i = 1, \dots, N$, then $\sum_{i=1}^N a_i g_i \in \mathcal{BF}$. Some examples of Bernstein functions are $g(x) = x^\alpha$, for $0 < \alpha < 1$, $g(x) = x/(1 + x)$ and $g(x) = \log(1 + x)$. The representation of the capacity function $\log(1 + x)$ in the form (4) is known as Frullani's integral [10, p. 6], and is given by

$$\log(1 + x) = \int_0^\infty (1 - e^{-sx}) \frac{e^{-s}}{s} \mathrm{d}s . \quad (5)$$

D. Thorin-Bernstein Functions

A Bernstein function g is called a Thorin-Bernstein function [9, pp. 73-79], if it admits the representation given by (4), where $s\mu(s)$ is c.m.. The family of all Thorin-Bernstein functions is

denoted by $\mathcal{TB}\mathcal{F}$. A necessary and sufficient condition for $g : (0, \infty) \rightarrow (0, \infty)$ to be in $\mathcal{TB}\mathcal{F}$ is that g can be represented as follows [9, p. 73]:

$$g(x) = a + bx + \int_{(0, \infty)} \log(1 + x/s) \mu(ds), \quad (6)$$

for some $a, b \geq 0$ and μ is a positive measure on $(0, \infty)$, which satisfies the convergence condition $\int_0^1 |\log s| \mu(ds) + \int_1^\infty s^{-1} \mu(ds) < \infty$. We refer to any $g_2 \in \mathcal{TB}\mathcal{F}$ which satisfies the property that $g_1(g_2(\cdot)) \in \mathcal{TB}\mathcal{F}$ whenever $g_1 \in \mathcal{TB}\mathcal{F}$ as a *composable* Thorin-Bernstein function (we denote the set of all such functions by $\mathcal{CTB}\mathcal{F}$). A sufficient condition for any g_2 to belong to $\mathcal{CTB}\mathcal{F}$ is that $g_2 \in \mathcal{TB}\mathcal{F}$ and $(dg_2(x)/dx)/g_2(x) \in \mathcal{S}$. Functions belonging to the class $\mathcal{TB}\mathcal{F}$ are of particular relevance to this paper, since the Shannon capacity function $C(x) := \log(1 + x)$ not only belongs to $\mathcal{B}\mathcal{F}$, but also belongs to $\mathcal{TB}\mathcal{F}$, as seen from (5) and (6).

It is useful to define a multivariate extension of a Thorin-Bernstein function. A function $g : \mathbb{R}^m \rightarrow \mathbb{R}$ belongs to $\mathcal{TB}\mathcal{F}_m$ if $g(x_1, \dots, x_m)$ is a Thorin-Bernstein function in each argument, when all other arguments are treated as constants. Further, if g is composable in each variable when all other variables are fixed, then g is said to belong to the set $\mathcal{CTB}\mathcal{F}_m$. An example of function in $\mathcal{CTB}\mathcal{F}_m$ can be verified to be $g(x_1, \dots, x_M) = \sum_{i=1}^M \alpha_i x_i, \alpha_i \geq 0, i = 1, \dots, M$.

E. Matrix Functions

Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$. If $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_N)$, we define $\phi(\mathbf{D}) = \text{diag}(\phi(\lambda_1), \dots, \phi(\lambda_N))$. If $\mathbf{A} \in \mathbb{S}_+^n$, so that $\mathbf{A} = \mathbf{U} \text{diag}(\lambda_1(\mathbf{A}), \dots, \lambda_N(\mathbf{A})) \mathbf{U}^H$, where \mathbf{U} is a unitary matrix, then we define $\phi(\mathbf{A}) = \mathbf{U} \phi(\mathbf{D}) \mathbf{U}^H$. In this way, $\phi(\mathbf{A})$ can be defined for all Hermitian matrices of any order [11]. In this work, the scalar function and its matrix extension are denoted using the same symbol, and the argument of the function defines the specific context. Matrix functions find applications in Section V.

F. Integral Stochastic Orders

Let \mathcal{G} denote a class of real valued functions $g : \mathbb{R}^+ \rightarrow \mathbb{R}$, and X and Y be random variables (RVs). We define the integral stochastic order with respect to \mathcal{G} as [6]:

$$X \leq_{\mathcal{G}} Y \Leftrightarrow \mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)] , \forall g \in \mathcal{G} . \quad (7)$$

In this case, \mathcal{G} is known as a generator of the order $\leq_{\mathcal{G}}$. We now give an example of an integral stochastic order relevant to this paper, by specifying the corresponding generator set of functions \mathcal{G} .

1) *Laplace Transform Order*: This partial order compares random variables based on their Laplace Transforms. Here, $\mathcal{G} = \{g(x) : g(x) = -\exp(-\rho x), \rho \geq 0\}$, so that $X \leq_{\text{Lt}} Y$ if and only if

$$\mathbb{E}[\exp(-\rho Y)] \leq \mathbb{E}[\exp(-\rho X)], \forall \rho \geq 0. \quad (8)$$

One useful property of LT ordered random variables is that for all c.m. functions g , we have

$$\mathbb{E}[g(Y)] \leq \mathbb{E}[g(X)]. \quad (9)$$

In other words, the generator \mathcal{G} can be enlarged to the set of all c.m. functions without changing the stochastic order [6]. Further, whenever $g \in \mathcal{BF}$, (9) holds with a reversal in the inequality. In a wireless communications context, let $\rho > 0$ be the average SNR, and $\rho X, \rho Y$ represent the instantaneous SNRs of two fading distributions. If $g(x)$ corresponds to the instantaneous symbol error rate $P_e(\rho x)$ of a modulation scheme with c.m. error rate function, then (9) can be used to obtain comparisons of averages of symbol error rates over pairs of fading channels, even in cases where a closed-form expression for the same is intractable.

Another useful result for Laplace transform ordered RVs is that, if X_1, \dots, X_M are independent and Y_1, \dots, Y_M are independent, and if $X_m \leq_{\text{Lt}} Y_m$, $m = 1, \dots, M$, then $g(X_1, \dots, X_M) \leq_{\text{Lt}} g(Y_1, \dots, Y_M)$, whenever $g(x_1, \dots, x_M)$ is a Bernstein function of x_i , when all other arguments of g are viewed as constants [9].

G. Shannon Transform

In what follows, we formally describe the Shannon transform, which is the basis of the proposed stochastic order in this paper. The Shannon transform of a non-negative random variable X is defined as [2, pp. 44]:

$$\overline{C}^{(X)}(\rho) := \mathbb{E}[\log(1 + \rho X)], \rho \geq 0. \quad (10)$$

Two new representations of $\overline{C}^{(X)}(\rho)$, which are useful in this paper are now obtained. Using (5), it is easy to show that (10) can be represented as a Laplace transform, given by

$$\overline{C}^{(X)}(\rho) = \int_0^\infty \exp(-u/\rho) \frac{1 - \phi_X(u)}{u} du, \quad (11)$$

for $\rho > 0$, where $\phi_X(u) := \mathbb{E}[\exp(-uX)]$, $u > 0$. Using (2) with (11), it is immediate that $\overline{C}^{(X)}(\rho)$ is a completely monotone function of $1/\rho$. A second representation of $\overline{C}^{(X)}(\rho)$ which can be derived from (11) shows that $\overline{C}^{(X)}(\rho)$ is also the Stieltjes transform [12, p. 325] of the complimentary CDF of X , when evaluated at $1/\rho$:

$$\overline{C}^{(X)}(\rho) = \int_0^\infty \frac{1 - F_X(u)}{(1/\rho + u)} du, \quad (12)$$

where $\rho > 0$. Representation (12) is used in proving some properties of the ergodic capacity order discussed in Section III-B. Additionally, (12) permits us to comment on the convergence of $\overline{C}^{(X)}(\rho)$:

Proposition 1. *If $\overline{C}^{(X)}(\rho)$ exists for any $\rho \in (0, \infty)$, then $\overline{C}^{(X)}(\rho)$ exists for every $\rho \in (0, \infty)$.*

Proof: From (12), it is seen that $\overline{C}^{(X)}(\rho)$ is the Stieltjes transform of a real valued function. If the Stieltjes transform of a function exists at any point on \mathbb{R}^+ , then it exists at all points on \mathbb{R}^+ [12, p. 326]. This completes the proof. ■

We now provide examples of random variables for which the ergodic capacity is finite for $\rho < \infty$ using the following proposition:

Proposition 2. *Let $F_X(\cdot)$ denote the cumulative distribution function of a RV X . If for some $\delta \in (0, 1]$, $\int_0^t 1 - F_X(u) du = O(t^{1-\delta})$, $t \rightarrow \infty$, then $\overline{C}^{(X)}(\rho) < \infty$.*

Proof: First, observe that $\int_0^\infty (s+t)^{-1} d\alpha(t)$ exists if $\alpha(t) = O(t^{1-\delta})$, $t \rightarrow \infty$, for some $\delta > 0$ [12, p. 330 (Theorem 3b)]. The proposition then follows by letting $\alpha(t) = \int_0^t 1 - F_X(u) du$. This completes the proof. ■

In Proposition 2, the case of $\delta = 1$ is equivalent to the condition that the mean of X is finite. It is therefore straightforward to see that the ergodic capacity of fading distributions such as Nakagami- m and Rician is finite at all finite SNR, since these distributions have finite average power. We now proceed to define a stochastic order for comparing fading distributions based on the Shannon transform.

III. THE ERGODIC CAPACITY ORDER

Recall that the ergodic capacity of a single-input single-output (SISO) system is given by $\mathbb{E}[\log(1 + \rho X)]$, where X is the square of the amplitude of the complex fading gain, and is defined as the instantaneous fading power of the channel. It is straightforward to see through an application of Jensen's inequality that the AWGN channel (with no fading) outperforms every fading distribution with same average channel power, in terms of the ergodic capacity at all SNR. However, given two fading distributions, it is not trivial to compare them based on the ergodic capacity, as obtaining a closed-form expression for the ergodic capacity of many fading channels is analytically intractable. Motivated by this, we propose a stochastic ordering method, which can be used to compare the ergodic capacity of two different fading channels.

A. Definition

Definition 1. If X and Y are arbitrary non-negative RVs, then X is said to be dominated by Y in the ergodic capacity order (i.e. $X \leq_c Y$), if and only if $\overline{C}^{(X)}(\rho) \leq \overline{C}^{(Y)}(\rho)$, $\forall \rho > 0$.

For this stochastic order, the generator is chosen as $\mathcal{G} = \{g(x) : g(x) = \log(1 + \rho x), \rho > 0\}$. In the situation that both these terms are infinite, we still say that $X \leq_c Y$. Distributions of interest for which the ergodic capacity is finite at all finite SNR can be determined using either Proposition 1 or Proposition 2. Next, some useful properties of the capacity order and a few examples of ergodic capacity ordered RVs are discussed.

B. Properties

The following properties hold for non-negative RVs.

S1: $X \leq_c Y \Leftrightarrow \mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$, $\forall g \in \mathcal{TB}\mathcal{F}$.

S2: $X \leq_c Y \Leftrightarrow g(X) \leq_c g(Y)$, $\forall g \in \mathcal{CTB}\mathcal{F}$.

S3: $X \leq_{\text{Lt}} Y \Rightarrow X \leq_c Y$.

S4: Let X_1, \dots, X_M independent and Y_1, \dots, Y_M independent. If $X_m \leq_c Y_m$, $m = 1, \dots, M$, then $g(X_1, \dots, X_M) \leq_c g(Y_1, \dots, Y_M)$, $\forall g \in \mathcal{CTB}\mathcal{F}_m$.

S5: If $X \leq_c Y$ and $Y \leq_c Z$, then $X \leq_c Z$.

S6: If $X \leq_c Y$ and $Y \leq_c X$, then $F_X(\cdot) = F_Y(\cdot)$ a.e..

The proofs of these properties can be found in Appendix A. A straightforward implication of Property S1 is that if $X \leq_c Y$, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$, since $g(x) = x$ is a Thorin-Bernstein function. In other words, if one fading channel has a higher ergodic capacity than another at all SNR, then it is necessary that the average fading power of the first channel is no smaller than that of the second. We further observe that Property S4 is true when $X_m \leq_{Lt} Y_m, m = 1, \dots, M$ with g being a \mathcal{BF} in each variable, while viewing all the other variables as constants [7]. Properties S5 and S6 together constitute the definition of a partial order, and consequently \leq_c is a partial order on non-negative RVs.

Interpreting ρX and ρY as the instantaneous SNRs of two different fading channels, Properties S1-S6 are useful in obtaining the conditions under which the ergodic capacity of a composite system with coding/decoding capabilities only at the transmitter/receiver under the channel Y is greater than that under X at all SNR. Although Property S3 suggests that every pair of Laplace transform ordered random variables also obey the ergodic capacity order, the converse is not true in general. A counterexample can be found in [1], [8]. Thus, it is possible that the average symbol error rate of differential binary phase shift keying modulation in channel X is less than that in Y at high SNR, while the situation reverses when the capacity achieving code is applied on both channels. Interpreting the ergodic capacity as what is achievable by coding over an i.i.d. time-extension of the channel, we reach the conclusion that even though Y offers more diversity than X for an uncoded system, the i.i.d. extension of X lends itself to more diversity than that of Y . To put it more simply, at high SNR, it is possible for one fading channel to be superior to another in terms of error rates in the absence of coding, while being inferior when the capacity achieving code is employed over both channels.

C. Examples

Next, we give examples of pairs of RVs X, Y relevant to wireless communications, for which $X \leq_c Y$ holds. In general, establishing capacity ordering using its definition is often inconclusive, since the corresponding integrals are intractable. Fortunately, using Property S3, it is possible to provide examples of pairs of RVs which obey capacity ordering. In what follows, examples of parametric fading distributions which obey the ergodic capacity order are given. These distributions are also known to satisfy the Laplace transform order [8].

1) *Nakagami Fading*: The Nakagami- m fading model, for which the envelope \sqrt{X} is Nakagami distributed, and the instantaneous fading power X is Gamma distributed, with PDF given by

$$f_X(x) = \frac{m^m}{\Gamma(m)} x^{m-1} \exp(-mx), x \geq 0, \quad (13)$$

where $m > 0$ is the line of sight parameter, and $\Gamma(r) := \int_0^\infty t^{r-1} \exp(-t) dt$ is the gamma function. If $X \sim \text{Gamma}(m^X)$, and $Y \sim \text{Gamma}(m^Y)$ with $m^X \leq m^Y$, then $X \leq_c Y$.

2) *Rician Fading*: The Rician fading model: In this case, the envelope of the fading i.e., \sqrt{X} is Rice distributed with line of sight parameter K , and the corresponding instantaneous fading power distribution is given by

$$f_X(x) = (K+1) \exp[-(K+1)x - K] I_0\left(2\sqrt{K(K+1)x}\right), \quad (14)$$

where $I_0(t) := \sum_{m=0}^\infty (t/2)^{2m} / (m! \Gamma(m+1))$ is the modified Bessel function of the first kind of order zero. If the distribution of X and Y have parameters K^X and K^Y respectively, with $K^X \leq K^Y$, then $X \leq_c Y$.

3) *Hoyt Fading*: The Nakagami- q (Hoyt) fading model: Here, the envelope of the fading RV, given by \sqrt{X} is Hoyt distributed, and the density of the (unit mean) instantaneous fading power is given by

$$f_X(x) = a \exp(-a^2 x) I_0(bx), \quad (15)$$

where $a = (1 + q^2)/2q$, $b = (1 - q^4)/4q^2$. If X and Y have parameters q^X and q^Y respectively, where $q^X \leq q^Y$, then $X \leq_c Y$.

For the cases of Nakagami, Rician and Hoyt fading, the increase in ergodic capacity with increase in the LoS parameter of the distribution is not due to an increase in the average fading power, since $\mathbb{E}[X] = \mathbb{E}[Y]$, which is independent of the LoS parameter.

In what follows, we show that ergodic capacity ordering of a given SISO system under two different fading channels can be used to make meaningful conclusions when a number of such systems are combined to form a system involving multiple random variables.

IV. SYSTEMS INVOLVING MULTIPLE RANDOM VARIABLES

In order to illustrate the applicability of the capacity order to compare the performance of systems, we provide examples of composite systems where capacity ordering of component SISO

systems can be used to conclude the capacity ordering of the system, and also some applications where this is not necessarily the case. Such generic conclusions can be made even when closed form expressions for the ergodic capacity are not available. Throughout, we assume that the receiver has a perfect estimate of the instantaneous fading power, while the transmitter does not possess any such information.

A. Diversity Combining Systems

As examples of systems involving multiple fading links, we first consider diversity combining schemes such as MRC and EGC using M receive antennas, for which we aim to compare the ergodic capacity under two different fading scenarios. Using the properties of the ergodic capacity order, we now show that diversity combining systems formed using a better set of components yields a system with a higher ergodic capacity, for the two schemes considered.

1) *Maximum Ratio Combining*: Conditioned on the instantaneous fading power $X_m = x_m$, $m = 1, \dots, M$, the fading power after combining is given by

$$g_{\text{MRC}}(x_1, \dots, x_M) = \sum_{m=1}^M x_m. \quad (16)$$

The ergodic capacity corresponding to this combining scheme is given by

$$\overline{C}_{\text{EGC}}^{(X)}(\rho) = \mathbb{E}[\log(1 + \rho g_{\text{MRC}}(X_1, \dots, X_M))] . \quad (17)$$

In order to compare the ergodic capacity of MRC in two different fading environments characterized by instantaneous fading powers (X_1, \dots, X_M) and (Y_1, \dots, Y_M) , we first verify that $g_{\text{MRC}}(\cdot)$ is a composable Thorin-Bernstein function. Then, we use Property S4 to conclude $\overline{C}_{\text{MRC}}^{(X)}(\rho) \leq \overline{C}_{\text{MRC}}^{(Y)}(\rho)$, $\forall \rho > 0$, when $X_m \leq_c Y_m$, $m = 1, \dots, M$.

In order to check if $g_{\text{MRC}}(\cdot) \in \mathcal{CTBF}$, we first show that $g \in \mathcal{TB}\mathcal{F}$. To this end, treat x_1 in $g_{\text{MRC}}(\cdot)$ as the variable, while treating other arguments as constants, to get $g_{\text{MRC}}(x_1; x_2, \dots, x_M) = x_1 + k$, where $k = \sum_{m=2}^M x_m$. Clearly, $g_{\text{MRC}}(x_1; x_2, \dots, x_M) \in \mathcal{TB}\mathcal{F}$, since it satisfies the conditions of (6) with $a = k$, $b = 1$ and $\mu(u) = 0$. Thus, by definition, $g_{\text{MRC}} \in \mathcal{TB}\mathcal{F}$. Next, g_{MRC} is shown to belong to \mathcal{CTBF} : $g'_{\text{MRC}}(x_1; x_2, \dots, x_M)/g_{\text{MRC}}(x_1; x_2, \dots, x_M) = (x_1 + k)^{-1}$, which is a Stieltjes function, as it satisfies (3) with $a = 0$, $b = 0$, and $\mu(s) = \delta(s)$.

Now, assuming $X_m \leq_c Y_m$, $m = 1, \dots, M$, we have from Property S4 $g_{\text{MRC}}(X_1, \dots, X_M) \leq_c g_{\text{MRC}}(Y_1, \dots, Y_M)$, which implies $\overline{C}_{\text{MRC}}^{(X)}(\rho) \leq \overline{C}_{\text{MRC}}^{(Y)}(\rho)$, $\forall \rho > 0$. Thus, if Y_m dominates X_m

in the capacity order for $m = 1, \dots, M$, then the MRC system with fading links given by Y_1, \dots, Y_M will have a higher ergodic capacity than that with X_1, \dots, X_M at all SNR.

2) *Equal Gain Combining*: For the case of equal gain combining, the ergodic capacity is given by

$$\overline{C}_{\text{EGC}}^{(X)}(\rho) = \mathbb{E} [\log (1 + \rho g_{\text{EGC}}(X_1, \dots, X_M))] , \quad (18)$$

where $g_{\text{EGC}}(\cdot)$ represents the combined instantaneous fading power, and is given by

$$g_{\text{EGC}}(x_1, \dots, x_M) = \left(\sum_{m=1}^M \sqrt{x_m} \right)^2 . \quad (19)$$

In order to verify that $g_{\text{EGC}} \in \mathcal{TB}\mathcal{F}$, treat x_1 as the variable and all the other arguments of g_{EGC} as constants, so that $g_{\text{EGC}}(x_1; x_2, \dots, x_M) = x_1 + 2\sqrt{x_1}k + k^2$, where $k = \sum_{m=2}^M x_m$. Clearly, $g_{\text{EGC}}(x_1; x_2, \dots, x_M) \in \mathcal{TB}\mathcal{F}$, since it satisfies the conditions of (6) with $a = k^2$, $b = 1$ and $\mu(s) = ks^{-1/2}/\pi$. Repeating the same argument with all other parameters of $g_{\text{EGC}}(\cdot)$, we conclude that $g_{\text{EGC}} \in \mathcal{TB}\mathcal{F}$. In addition, $g_{\text{EGC}}(\cdot)$ is composable, as seen from the following arguments. Let $h(x) := g'_{\text{EGC}}(x; k)/g_{\text{EGC}}(x; k) = (x + k\sqrt{x_1})^{-1}$. To show that $h \in \mathcal{S}$, observe that $(h(x^{-1}))^{-1} = x^{-1} + kx^{-1/2}$ is a Stieltjes function, since any function of the form $x^{\alpha-1}$, $0 \leq \alpha \leq 1$ is a Stieltjes function [9, p. 13], and positive linear combinations of Stieltjes functions also yields a Stieltjes function. To complete the argument, since $(h(x^{-1}))^{-1} \in \mathcal{S}$, $h(x)$ must also belong to \mathcal{S} [9, p. 66]. Consequently, $g_{\text{EGC}}(\cdot) \in \mathcal{CTB}\mathcal{F}$.

Now, following similar arguments as in the MRC case, we infer that if a collection of SISO systems with higher ergodic capacity is combined to form an EGC system, then the composite EGC system will have higher overall ergodic capacity.

B. Multi-Hop Amplify-Forward Relay System

We now turn our attention to multi-hop amplify-forward (AF) relay systems. For this system, even though a set of better channels (in the ergodic capacity order) is combined to form the relay, it need not necessarily result in a relay with a higher ergodic capacity, as described below. Consider an M -hop AF relay (refer Figure 1), which is described as follows:

$$y_m = \sqrt{\alpha_m \rho} h_m s + v_m , m = 1, \dots, M , \quad (20)$$

where $y_m, m = 1, \dots, M - 1$ is the received signal at the m^{th} half-duplex variable-gain relay, h_m is the complex ergodic fading on the m^{th} hop, which is i.i.d across time, s is the transmitted

symbol, ρ is the average SNR, v_m is the AWGN which is i.i.d across hops as well as time, and y_M is the signal received at the destination. It is assumed that coding/decoding capabilities are provided to the transmitter/receiver alone, and the m^{th} relay, which has knowledge of X_m , merely amplifies and forwards the signal using a variable gain given by $\alpha_m = \rho/(\rho X_m + 1)$ [13], where $X_m := |h_m|^2$ is the instantaneous fading power of the m^{th} hop, for $m = 1, \dots, M-1$. In this case, the ergodic capacity is given by

$$\overline{C}_{\text{MH-AF}}^{(X)}(\rho) = \mathbb{E} [\log (1 + \rho g_{\text{MH-AF}}(X_1, \dots, X_M))] , \quad (21)$$

where $g_{\text{MH-AF}}(x_1, \dots, x_M) := (\prod_{m=1}^M [(1 + x_m^{-1})] - 1)^{-1}$. Exact expressions for the ergodic capacity in arbitrary fading channels are intractable, even for the two-hop case. Previously, the ergodic capacity of such a relay in fading channels has been obtained as an infinite series in [14].

In order to compare the performance of the AF relay in two different fading scenarios, let X_m and Y_m denote the instantaneous fading power of the m^{th} link of the first and second fading channels respectively, for $m = 1, \dots, M$. Further, assume that $X_m \leq_c Y_m$, $m = 1, \dots, M$. In order to use Property S4 to obtain conclusions about the comparative ergodic capacities, we first check if $g_{\text{MH-AF}} \in \mathcal{CTBF}$. To this end, we first check if $g_{\text{MH-AF}} \in \mathcal{TF}$. Treating x_1 as the variable and all other arguments of $g_{\text{MH-AF}}(\cdot)$ as constants, we get $g_{\text{MH-AF}}(x_1; x_2, \dots, x_M) = kx_1/(x_1 + k)$, where $k = g_{\text{MH-AF}}(x_2, \dots, x_M)$. From [9, pp. 218], it is seen that $g_{\text{MH-AF}} \notin \mathcal{TF}$. Therefore, $g \notin \mathcal{CTBF}$, and Property S4 cannot be used to conclude anything about the ergodic capacity of X versus that of Y . Nevertheless, by straight-forward differentiation with respect to x_i , it is observed that $g_{\text{MH-AF}}(\cdot)$ is a Bernstein function of x_i . Therefore, $g_{\text{MH-AF}} \in \mathcal{BF}$, but $g_{\text{MH-AF}} \notin \mathcal{TF}$. As a result, if the instantaneous fading powers also satisfy the stronger criterion $X_m \leq_{\text{Lt}} Y_m$, $m = 1, \dots, M$, then we get $g_{\text{MH-AF}}(X_1, \dots, X_M) \leq_{\text{Lt}} g_{\text{MH-AF}}(Y_1, \dots, Y_M)$, and therefore $g_{\text{MH-AF}}(X_1, \dots, X_M) \leq_c g_{\text{MH-AF}}(Y_1, \dots, Y_M)$. However, if $X_m \leq_c Y_m$, while $X_m \not\leq_{\text{Lt}} Y_m$, $m = 1, \dots, M$ then we cannot conclude $g_{\text{MH-AF}}(X_1, \dots, X_M) \leq_c g_{\text{MH-AF}}(Y_1, \dots, Y_M)$. The latter is observed in the case of an interference dominated channel, where the instantaneous fading power to interference power ratio X_m are independent and Pareto distributed with parameter β_X :

$$F_{X_m}(x) = \frac{x^{\beta_X}}{1 + x^{\beta_X}}, x > 0, \beta_X > 0 , \quad (22)$$

and Y_m similarly with parameter β_Y , where $\beta_X \leq \beta_Y$. For example, Fig. 2 illustrates the numerically evaluated ergodic capacities of a multi-hop relay with $M = 3$ hops under Pareto

distributed signal-to-interference ratio with parameters $\beta_X = 1$ and $\beta_Y = 3$, so that for each hop $X_m \leq_c Y_m, m = 1, 2, 3$ is satisfied. It is observed from Fig. 2 that for $\rho < \rho_0$, where $\rho_0 \approx 5$ dB, X is a better channel than Y in the capacity order, while for $\rho \geq \rho_0$, the situation is reversed.

C. Fading Multiple Access Channel

In this example, we focus on comparing the ergodic capacity regions of a multi-user Gaussian MAC network in two different fading scenarios. Consider the following system model:

$$y = \sqrt{\rho} \sum_{m=1}^M h_m s_m + v, \quad (23)$$

where ρ is the average SNR of each user, s_m is the transmitted symbol of user m , h_m is the complex i.i.d (across time) ergodic fading between each user and the destination, and v is the AWGN at the receiver. It is assumed that only the receiver possesses CSI of all the users. The receiver intends to decode the signals from all the users. If $X_m := |h_m|^2, m = 1, \dots, M$, then the ergodic capacity region is the set of all rate M -tuples that satisfy [15, pp. 407],

$$\sum_{m \in \mathcal{S}} R_m(\rho) \leq \mathbb{E} \left[\log \left(1 + \rho \sum_{\mathcal{S}} X_m \right) \right], \quad (24)$$

where $\mathcal{S} \subset 2^{\{1, \dots, M\}}$. Let $g_{\text{MAC}, \mathcal{S}}(X_1, \dots, X_M) := \sum_{\mathcal{S}} X_m$. Clearly, $g_{\text{MAC}, \mathcal{S}}(X_1, \dots, X_M) \in \mathcal{CTBF}_{\text{card } \mathcal{S}}$, if $X_m \leq_c Y_m, m = 1, \dots, M$, from Property S4 it follows that

$$g_{\text{MAC}, \mathcal{S}}(X_1, \dots, X_M) \leq_c g_{\text{MAC}, \mathcal{S}}(Y_1, \dots, Y_M), \forall \mathcal{S} \subset 2^{\{1, \dots, M\}}. \quad (25)$$

Thus, if each user of the system X has a higher ergodic capacity than the corresponding user in the system Y , then $\overline{C}_{\text{MAC}}^{(X)}(\rho) \subseteq \overline{C}_{\text{MAC}}^{(Y)}(\rho), \forall \rho > 0$.

V. MIMO ERGODIC CAPACITY ORDER

In this section, the ergodic capacity ordering of MIMO systems is presented. Some properties of this stochastic order are discussed, and an application of this framework in a MIMO MAC setting is presented. Before doing so, we formally define a MIMO system through its single letter characterization:

$$\mathbf{y} = \sqrt{\rho} \mathbf{H} \mathbf{s} + \mathbf{v}, \quad (26)$$

where \mathbf{H} is a complex $N_R \times N_T$ random matrix which captures the effect of ergodic quasi-static fading, $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the additive noise, \mathbf{s} is the transmitted symbol vector, and ρ is the

average SNR per transmit antenna. \mathbf{H} and \mathbf{v} are assumed to be i.i.d across time, as a result of which a time index has not been used in (26). Further, it is assumed that the receiver tracks the channel fading realizations \mathbf{H} , while no such CSI is available at the transmitter. For this system model, the ergodic capacity is given by $\overline{C}_{\text{MIMO}}^{(X)}(\rho) = \mathbb{E} [\log \det (\mathbf{I} + \rho \mathbf{X})]$, where \mathbf{X} denotes the instantaneous fading power, which equals $\mathbf{H}^H \mathbf{H}$. In what follows, we define a partial order on the instantaneous fading power, which can be used to compare the ergodic capacity of composite MIMO systems under two different fading environments.

A. Definition and Properties

Definition 2. If two random matrices \mathbf{X}, \mathbf{Y} belong to \mathbb{S}_+^n , then $\mathbf{X} \preceq_c \mathbf{Y}$ if and only if $\mathbf{X} \preceq_c \mathbf{Y} \Leftrightarrow \mathbb{E} [\text{tr} \log (\mathbf{I} + \rho \mathbf{X})] \leq \mathbb{E} [\text{tr} \log (\mathbf{I} + \rho \mathbf{Y})], \forall \rho > 0$.

In Definition 2, $\log(\cdot)$ is to be viewed as a matrix function, in the sense of Section II-E. It is easy to show that $\mathbf{X} \preceq_c \mathbf{Y}$ is equivalent to $\mathbb{E} [\log \det (\mathbf{I} + \rho \mathbf{X})] \leq \mathbb{E} [\log \det (\mathbf{I} + \rho \mathbf{Y})], \forall \rho > 0$. In contrast to the ergodic capacity order on random variables, the MIMO ergodic capacity corresponding to two different random matrices \mathbf{X} and \mathbf{Y} may be identical (for example, when $\mathbf{Y} = \mathbf{U} \mathbf{X} \mathbf{U}^H$, where \mathbf{U} is a unitary matrix). In this circumstance, we write $\mathbf{X} =_c \mathbf{Y}$. In what follows, some properties of the MIMO ergodic capacity order are developed, which can be viewed as matrix analogues to the properties developed in Section III-B. The following properties are true for positive semi-definite symmetric random matrices.

M1: If $\mathbf{X}, \mathbf{Y} \in \mathbb{S}_+^n$, then $\mathbf{X} \preceq_c \mathbf{Y} \Leftrightarrow \mathbb{E} [\text{tr} g(\mathbf{X})] \leq \mathbb{E} [\text{tr} g(\mathbf{Y})]$, for all $g : \mathbb{R} \rightarrow \mathbb{R}$, such that $g \in \mathcal{TB}\mathcal{F}$.

M2: If $\mathbf{X}, \mathbf{Y} \in \mathbb{S}_+^n$, then $\mathbf{X} \preceq_c \mathbf{Y} \Leftrightarrow g(\mathbf{X}) \preceq_c g(\mathbf{Y})$, for all $g : \mathbb{R} \rightarrow \mathbb{R}$, such that $g \in \mathcal{CTB}\mathcal{F}$.

M3: If $\mathbf{X}, \mathbf{Y} \in \mathbb{S}_+^n$ and $\mathbb{E} [\text{tr} \exp(-\rho \mathbf{X})] \geq \mathbb{E} [\text{tr} \exp(-\rho \mathbf{Y})] \forall \rho \geq 0$ then $\mathbf{X} \preceq_c \mathbf{Y}$.

M4: Let $\mathbf{X}_{1:M} := [\mathbf{X}_1, \dots, \mathbf{X}_M]$, $\mathbf{Y}_{1:M} := [\mathbf{Y}_1, \dots, \mathbf{Y}_M]$, where $\mathbf{X}_i \in \mathbb{S}_+^n$ are independent random matrices, for $i = 1, \dots, M$, and likewise for \mathbf{Y}_i . Let $g(\mathbf{X}_{1:M}) := g(\mathbf{X}_1, \dots, \mathbf{X}_M)$, i.e., g operates on M \mathbb{S}_+^n matrices and produces a \mathbb{S}_+^n matrix. If $g : \mathbb{R}^M \rightarrow \mathbb{R}$ is such that $g \in \mathcal{CTB}\mathcal{F}_m$ then $g(\mathbf{X}_{1:M}) \preceq_c g(\mathbf{Y}_{1:M})$.

M5: If $\mathbf{X} \preceq_c \mathbf{Y}$, and $\mathbf{Y} \preceq_c \mathbf{Z}$, then $\mathbf{X} \preceq_c \mathbf{Z}$.

M6: $\mathbf{X} =_c \mathbf{Y}$ if and only if $\sum_{i=1}^n F_{\lambda_i(\mathbf{X})}(u) = \sum_{i=1}^n F_{\lambda_i(\mathbf{Y})}(u)$, where $F_{\lambda_i(\mathbf{X})}(\cdot)$ is the marginal CDF of the i^{th} largest eigenvalue of \mathbf{X} .

The proofs of properties M1-M4, and M6 can be found in Appendix B, while Property M5 is straight-forward to establish, and its proof is omitted. Property M3 provides a useful sufficient condition to verify if two random matrices obey the MIMO ergodic capacity order. This is because $\mathbb{E}[\text{tr} \exp(-\rho \mathbf{X})] \geq \mathbb{E}[\text{tr} \exp(-\rho \mathbf{Y})] \quad \forall \rho \geq 0$ is equivalent to $\sum_{i=1}^n \mathbb{E}[\exp(-\rho \lambda_i(\mathbf{X}))] \geq \sum_{i=1}^n \mathbb{E}[\exp(-\rho \lambda_i(\mathbf{Y}))], \forall \rho \geq 0$, and Laplace transforms of the eigenvalue distributions are more easy to compute, when compared to the expectations of the log-determinants. Next, we form an interesting interpretation of Property M6. From Property M6, it follows that $\mathbf{X} =_c \mathbf{Y}$ if and only if $\mathbb{E}_Q[F_{\lambda_Q(\mathbf{X})}(u)] = \mathbb{E}_Q[F_{\lambda_Q(\mathbf{Y})}(u)]$, where Q is a uniformly distributed integer over $[1, n]$. In other words, if the distribution of an eigenvalue picked randomly and uniformly from both matrices is identical, then the two random matrices are regarded to be the same with respect to the MIMO ergodic capacity order.

Although the proposed definition of the MIMO ergodic capacity order is one of many different possible partial orders on matrices, we assert that it is a natural generalization of the ergodic capacity order defined in Section III. This is also elucidated by the fact that the properties M1-M5 are indeed straight-forward matrix generalizations of properties S1-S5. Further, the MIMO ergodic capacity order bears the following connection with the ergodic capacity order defined for random variables:

Theorem 1. *Let $\lambda_Q(\mathbf{X}) \leq_c \lambda_Q(\mathbf{Y})$, where $\lambda_Q(\mathbf{X})$ is an eigenvalue of \mathbf{X} picked uniformly from the set of eigenvalues of \mathbf{X} . Then $\mathbf{X} \preceq_c \mathbf{Y}$. Conversely, if $\mathbf{X} \preceq_c \mathbf{Y}$, then $\lambda_Q(\mathbf{X}) \leq_c \lambda_Q(\mathbf{Y})$.*

Proof: See Appendix C. ■

Given two MIMO fading systems \mathbf{X} and \mathbf{Y} , Theorem 1 implies that \mathbf{Y} dominates \mathbf{X} in the MIMO ergodic capacity order, if and only if a uniformly randomly selected eigen-channel of \mathbf{Y} has a larger ergodic capacity than that of a uniformly randomly selected eigen-channel of \mathbf{X} .

An illustrative example to elucidate the efficacy of the MIMO ergodic capacity order is the M user Gaussian MIMO MAC, where user i possesses N_t antennas. We assume that only the receiver has CSI, and that each antenna of each user transmits independent signals. Further, each user is allocated the same transmit power ρ per transmit antenna. In this case, the ergodic

capacity region $\mathcal{C}_{\text{MIMO-MAC}}(\rho)$ is given by [16]:

$$\mathcal{C}_{\text{MIMO-MAC}}^X(\rho) := \left\{ (R_1, \dots, R_M) : \sum_{i \in \mathcal{S}} R_i \leq \mathbb{E} [\log \det (\mathbf{I} + \rho g_{\text{MIMO-MAC}, \mathcal{S}}(\mathbf{X}_{1:M}))] \right\},$$

$$\forall \mathcal{S} \subseteq \{1, \dots, M\}, \quad (27)$$

where $g_{\text{MIMO-MAC}, \mathcal{S}}(\mathbf{X}_{1:M}) := \sum_{i \in \mathcal{S}} \mathbf{X}_i$, with $\mathcal{S} \subseteq \{1, \dots, M\}$. Clearly, when \mathbf{X}_i is assumed to be the variable while viewing all other arguments of $g_{\text{MIMO-MAC}, \mathcal{S}}(\cdot)$ as constant matrices, it can be seen that $g_{\text{MIMO-MAC}, \mathcal{S}}(\cdot)$ is a Thorin-Bernstein matrix function of \mathbf{X}_i , for $i = 1, \dots, M$. Therefore, through property M4, $g_{\text{MIMO-MAC}, \mathcal{S}}(\mathbf{X}_{1:M}) \preceq_c g_{\text{MIMO-MAC}, \mathcal{S}}(\mathbf{Y}_{1:M})$, whenever $\mathbf{X}_i \preceq_c \mathbf{Y}_i, i = 1, \dots, M$. Consequentially, $\mathcal{C}_{\text{MIMO-MAC}}^X(\rho) \subseteq \mathcal{C}_{\text{MIMO-MAC}}^Y(\rho), \forall \rho > 0$.

VI. CONCLUSION

The ergodic capacity order and its properties can be exploited to obtain comparisons of ergodic capacities of composite systems across two different fading channels whose instantaneous SNRs satisfy the ergodic capacity order. For systems such as MRC and EGC which involve multiple instantaneous SNR RVs, we conclude that combining a better set of channels (in the ergodic capacity order) produces a system with a higher ergodic capacity. This conclusion is true for all systems whose end-to-end instantaneous SNR belongs to the \mathcal{CTBF}_m set. For systems whose end-to-end SNR does not belong to \mathcal{CTBF}_m , component-wise ergodic capacity ordering of instantaneous SNR need not produce a system with a higher ergodic capacity. An example to illustrate this point is the AF relay for which the instantaneous SINR is Pareto distributed. An extension of the ergodic capacity order to MIMO systems is also proposed herein. The properties of the ergodic capacity order can be used to compare the capacity regions of systems such as the multi-user MAC in two different fading environments, for both the single and multiple antenna case.

APPENDIX A

PROOFS: PROPERTIES OF ERGODIC CAPACITY ORDER

Proof of Property S1

Let X and Y be non-negative RVs, and $g \in \mathcal{TB}\mathcal{F}$. Using (6), we can write

$$\mathbb{E} [g(X)] = \mathbb{E} \left[a + bX + \int_0^\infty \log \left(1 + \frac{X}{s} \right) \mu(ds) \right], \quad (28)$$

for some $a, b \geq 0$ and $\mu(s) \geq 0$, which satisfies the convergence conditions following (6). Commuting the expectation and integral in (28), we get

$$\mathbb{E}[g(X)] = a + b\mathbb{E}[X] + \int_0^\infty \mathbb{E}\left[\log\left(1 + \frac{X}{s}\right)\right] \mu(ds) \quad (29)$$

$$\leq a + b\mathbb{E}[Y] + \int_0^\infty \mathbb{E}\left[\log\left(1 + \frac{Y}{s}\right)\right] \mu(ds) \quad (30)$$

$$= \mathbb{E}[g(Y)] ,$$

where (30) follows from (29) and the fact that $X \leq_c Y \Rightarrow \mathbb{E}[X] \leq \mathbb{E}[Y]$. To prove $X \leq_c Y \Rightarrow \mathbb{E}[X] \leq \mathbb{E}[Y]$, for $\rho > 0$, we have $(1/\rho)\mathbb{E}[\log(1 + \rho X)] \leq (1/\rho)\mathbb{E}[\log(1 + \rho Y)]$. Now, moving the $1/\rho$ term inside the expectation on both sides, we take the limit as $\rho \rightarrow 0$. The limit and expectation can be commuted, since $\mathbb{E}[\log(1 + \rho X)] \leq \mathbb{E}[X] \forall \rho > 0$, and consequently, the dominated convergence criterion is satisfied. This result can also be shown to hold for the case when $\mathbb{E}[Y] = \infty$, by using the relation $\mathbb{E}[\log(1 + Y)] \leq \mathbb{E}[Y]$.

To prove the converse, we begin with $\mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ for $g \in \mathcal{TB}\mathcal{F}$. Choosing $g(x) = \log(1 + x)$, which is $\mathcal{TB}\mathcal{F}$, we get $X \leq_c Y$. This concludes the proof.

Proof of Property S2

Let X, Y be non-negative RVs, and $g, \phi : \mathbb{R}^+ \rightarrow \mathbb{R} \in \mathcal{TB}\mathcal{F}$. Since $\phi \circ g \in \mathcal{TB}\mathcal{F}$ if and only if $g \in \mathcal{CTB}\mathcal{F}$, from Property S1, if $X \leq_c Y$, we have $\mathbb{E}[\phi(g(X))] \leq \mathbb{E}[\phi(g(Y))]$. Since $\phi \in \mathcal{TB}\mathcal{F}$, this implies $g(X) \leq_c g(Y)$.

To prove the converse, consider $g(X) \leq_c g(Y)$ for $g \in \mathcal{CTB}\mathcal{F}$. Choosing $g(x) = x$, we get $X \leq_c Y$, which completes the proof.

Proof of Property S3

Recall from Section II-F1 that $X \leq_{\text{Lt}} Y \Leftrightarrow \mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ for all $g \in \mathcal{B}\mathcal{F}$. The property then follows by observing that $g(x) = \log(1 + x)$ is a Bernstein function.

Proof of Property S4

This property is proved using mathematical induction. To begin with, let $\phi : \mathbb{R} \rightarrow \mathbb{R}^+ \in \mathcal{TB}\mathcal{F}$, and $\mathbf{x}_{1:m} := [X_1, \dots, X_m]$ have independent and non-negative RVs as components. Assume

likewise for $\mathbf{y}_{1:m} := [Y_1, \dots, Y_m]$. Now, for $m = 1$, Property S4 is true due to Property S2. Next, let us assume Property S4 to be true for vectors of length $m - 1$. Thus, for any $g \in \mathcal{CTBF}_m$, it is straightforward to see that $g([a \ \mathbf{x}_{1:m-1}]) \leq_c g([a \ \mathbf{y}_{1:m-1}])$, where $g([a \ \mathbf{x}_{1:m-1}]) := g(a, X_1, \dots, X_{m-1})$, where $a \in \mathbb{R}^+$. This implies

$$\mathbb{E} [\phi (g([a \ \mathbf{x}_{1:m-1}]))] \leq \mathbb{E} [\phi (g([a \ \mathbf{y}_{1:m-1}]))] . \quad (31)$$

Next, for vectors of length m , consider

$$\mathbb{E} [\phi (g(\mathbf{x}_{1:m})) | X_1 = x] = \mathbb{E} [\phi (g([x \ \mathbf{x}_{2:m}]))] \quad (32)$$

$$\leq \mathbb{E} [\phi (g([x \ \mathbf{y}_{2:m}]))] = \mathbb{E} [\phi (g(\mathbf{y}_{1:m})) | Y_1 = x] , \quad (33)$$

where (33) follows from (32) due to (31). Now, taking the expectation with respect to X_1 on the left hand side of (32) and the right hand side of (33), we get $\mathbb{E} [\phi (g(X_1, \dots, X_m))] \leq \mathbb{E} [\phi (g(Y_1, \dots, Y_m))]$. Equations (32) and (33) can be repeated by conditioning on all the other RVs to complete the proof.

Proof of Property S5

Assume $X \leq_c Y$, and $Y \leq_c Z$. By definition, this implies $E \log (1 + \rho X) \leq E \log (1 + \rho Y)$, $\forall \rho \geq 0$, and $E \log (1 + \rho Y) \leq E \log (1 + \rho Z)$, $\forall \rho \geq 0$. Therefore, $E \log (1 + \rho X) \leq E \log (1 + \rho Z)$, $\forall \rho \geq 0$, which implies $X \leq_c Z$.

Proof of Property S6

To establish Property S6, it is sufficient to show that the Shannon transform of a RV has a one-to-one correspondence with the distribution of the RV. To this end, recall from (12) that for any non-negative RV X , $\overline{C}^{(X)}(\rho)$ can be written as the Stieltjes transform of $1 - F_X(\cdot)$. Since the Stieltjes transform of a function with bounded variation is unique [12, p. 336], and $1 - F_X(\cdot)$ is of bounded variation, the property follows.

APPENDIX B

PROOFS: PROPERTIES OF MIMO ERGODIC CAPACITY ORDER

Proof of Property M1

Assume $\mathbf{X} \preceq_c \mathbf{Y}$. Using the identity $\det(\mathbf{I} + \rho\mathbf{X}) = \prod_{i=1}^n (1 + \rho\lambda_i(\mathbf{X}))$, we can write

$$\mathbf{X} \preceq_c \mathbf{Y} \Leftrightarrow \mathbb{E} \left[\sum_{i=1}^n \log(1 + \rho\lambda_i(\mathbf{X})) \right] \leq \mathbb{E} \left[\sum_{i=1}^n \log(1 + \rho\lambda_i(\mathbf{Y})) \right], \quad (34)$$

$\forall \rho > 0$. It now follows that

$$\mathbf{X} \preceq_c \mathbf{Y} \Leftrightarrow \mathbb{E} \left[\sum_{i=1}^n \log(1 + t\rho\lambda_i(\mathbf{X}))\mu(t) \right] \leq \mathbb{E} \left[\sum_{i=1}^n \log(1 + t\rho\lambda_i(\mathbf{Y}))\mu(t) \right], \quad (35)$$

for all $\mu(t) \geq 0$, $\rho > 0$ and $t > 0$. Integrating the right hand side of (35) over t in the interval $[0, \infty)$ preserves the inequality in (35). Therefore,

$$\mathbf{X} \preceq_c \mathbf{Y} \Rightarrow \mathbb{E} \left[\sum_{i=1}^n \int_0^\infty \log(1 + t\rho\lambda_i(\mathbf{X}))\mu(dt) \right] \leq \mathbb{E} \left[\sum_{i=1}^n \int_0^\infty \log(1 + t\rho\lambda_i(\mathbf{Y}))\mu(dt) \right], \quad (36)$$

$\forall \rho > 0$. The summand in (36) is an arbitrary Thorin-Bernstein function, since μ is an arbitrary non-negative function. Denoting this Thorin-Bernstein function by g , the direct part of the property is proved by observing from Section II-E that $\mathbb{E}[\sum_{i=1}^n g(\lambda_i(\mathbf{X}))] = \mathbb{E}[\text{tr } g(\mathbf{X})]$. To prove the converse, choose $g(\mathbf{A}) = \log(\mathbf{I} + \rho\mathbf{A})$.

Proof of Property M2

Let $\mathbf{X}, \mathbf{Y} \in \mathbb{S}_+^n$, and $\mathbf{X} \preceq_c \mathbf{Y}$. Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$ belong to $\mathcal{TB}\mathcal{F}$, and $g : \mathbb{R} \rightarrow \mathbb{R}$ belong to $\mathcal{CTB}\mathcal{F}$. Using the definition of matrix functions, it is easy to see that $f(\mathbf{X}) := \phi(g(\mathbf{X})) \in \mathcal{TB}\mathcal{F}$. From Property M1, it is seen that $\mathbf{X} \preceq_c \mathbf{Y} \Leftrightarrow \mathbb{E}[\text{tr } \phi(g(\mathbf{X}))] \leq \mathbb{E}[\text{tr } \phi(g(\mathbf{Y}))]$. In other words, $g(\mathbf{X}) \preceq_c g(\mathbf{Y})$, which proves the direct part of the property. To see the converse, choose f as the identity map.

Proof of Property M3

Let $\mathbf{X}, \mathbf{Y} \in \mathbb{S}_+^n$, and $\mathbf{X} \preceq_c \mathbf{Y}$. Using Frullani's formula (5), it is evident that an equivalent condition to $\mathbf{X} \preceq_c \mathbf{Y}$ is given by

$$\mathbf{X} \preceq_c \mathbf{Y} \Leftrightarrow \mathbb{E} \left[\int_0^\infty \frac{e^{-s}}{s} \sum_{i=1}^n (1 - \exp(-\rho s \lambda_i(\mathbf{X}))) ds \right] \leq \mathbb{E} \left[\int_0^\infty \frac{e^{-s}}{s} \sum_{i=1}^n (1 - \exp(-\rho s \lambda_i(\mathbf{Y}))) ds \right]. \quad (37)$$

Commuting the expectation and integral in (37), we get

$$\mathbf{X} \preceq_c \mathbf{Y} \Leftrightarrow \int_0^\infty \frac{e^{-s}}{s} \mathbb{E} \left[\sum_{i=1}^n \exp(-\rho s \lambda_i(\mathbf{X})) \right] ds \geq \int_0^\infty \frac{e^{-s}}{s} \mathbb{E} \left[\sum_{i=1}^n \exp(-\rho s \lambda_i(\mathbf{Y})) \right] ds. \quad (38)$$

Therefore, if $\mathbb{E} [\sum_{i=1}^n \exp(-\rho \lambda_i(\mathbf{X}))] \geq \mathbb{E} [\sum_{i=1}^n \exp(-\rho \lambda_i(\mathbf{Y}))]$, $\rho > 0$, then $\mathbf{X} \preceq_c \mathbf{Y}$. The property then follows through the observation that $\mathbb{E} [\sum_{i=1}^n \exp(-\rho \lambda_i(\mathbf{X}))] = \mathbb{E} [\text{tr} \exp(-\rho \mathbf{X})]$.

Proof of Property M4

This property is proved using mathematical induction. To begin with, choose a matrix function $\phi \in \mathcal{TB}\mathcal{F}$, and $\mathbf{X}_{1:m} := [\mathbf{X}_1, \dots, \mathbf{X}_m]$ have independent and non-negative random matrices as components. Assume likewise for $\mathbf{Y}_{1:m} := [\mathbf{Y}_1, \dots, \mathbf{Y}_m]$. Now, for $m = 1$, Property M4 is true due to Property M1. Next, let us assume Property M4 to be true for sequences of length $m - 1$. Thus, for $g \in \mathcal{CTB}\mathcal{F}_m$ we have $g([\mathbf{C} \mathbf{X}_{1:m-1}]) \preceq_c g([\mathbf{C} \mathbf{Y}_{1:m-1}])$, where $g([\mathbf{C} \mathbf{X}_{1:m-1}]) := g(\mathbf{C}, \mathbf{X}_1, \dots, \mathbf{X}_{m-1})$, and $\mathbf{C} \in \mathbb{S}_+^n$. This implies

$$\mathbb{E} [\text{tr} \phi (g([\mathbf{C} \mathbf{X}_{1:m-1}]))] \leq \mathbb{E} [\text{tr} \phi (g([\mathbf{C} \mathbf{Y}_{1:m-1}]))] . \quad (39)$$

Next, for sequences of length m , consider

$$\mathbb{E} [\text{tr} \phi (g(\mathbf{X}_{1:m})) | \mathbf{X}_1 = \mathbf{A}] = \mathbb{E} [\text{tr} \phi (g([\mathbf{A} \mathbf{X}_{2:m}]))] \quad (40)$$

$$\leq \mathbb{E} [\text{tr} \phi (g([\mathbf{A} \mathbf{Y}_{2:m}]))] = \mathbb{E} [\text{tr} \phi (g(\mathbf{Y}_{1:m})) | \mathbf{Y}_1 = \mathbf{B}] , \quad (41)$$

where (41) follows from (40) due to (39). Now, taking the expectation with respect to \mathbf{X}_1 on the left hand side of (40) and the right hand side of (41), we get $\mathbb{E} [\text{tr} \phi (g(\mathbf{X}_1, \dots, \mathbf{X}_m))] \leq \mathbb{E} [\text{tr} \phi (g(\mathbf{Y}_1, \dots, \mathbf{Y}_m))]$. Equations (40) and (41) can be repeated by conditioning on all the other random matrices to complete the proof.

Proof of Property M6

To prove this property, let $\mathbf{X}, \mathbf{Y} \in \mathbb{S}_+^n$, and $\mathbb{E} [\log \det (\mathbf{I} + \rho \mathbf{X})] = \mathbb{E} [\log \det (\mathbf{I} + \rho \mathbf{Y})]$. Using the representation of the log-determinant in terms of the eigenvalues, and (12), it is seen that

$$\mathbb{E} [\log \det (\mathbf{I} + \rho \mathbf{X})] = \mathbb{E} [\log \det (\mathbf{I} + \rho \mathbf{Y})] \Leftrightarrow \int_0^\infty \frac{\sum_{i=1}^n 1 - F_{\lambda_i(\mathbf{X})}(u)}{1/\rho + u} du = \int_0^\infty \frac{\sum_{i=1}^n 1 - F_{\lambda_i(\mathbf{Y})}(u)}{1/\rho + u} du . \quad (42)$$

To see the direct part of the Property, recall the Stieltjes transform of a function of bounded variation is in a one-to-one correspondence with the function, and $\sum_{i=1}^n 1 - F_{\lambda_i(\mathbf{X})}(u)$ is of bounded variation. It is therefore immediate that if $\mathbb{E}[\log \det(\mathbf{I} + \rho \mathbf{X})] = \mathbb{E}[\log \det(\mathbf{I} + \rho \mathbf{Y})]$, then $\sum_{i=1}^n F_{\lambda_i(\mathbf{X})}(u) = \sum_{i=1}^n F_{\lambda_i(\mathbf{Y})}(u)$, a.e.. To prove the converse, assume $\sum_{i=1}^n F_{\lambda_i(\mathbf{X})}(u) = \sum_{i=1}^n F_{\lambda_i(\mathbf{Y})}(u)$ a.e.. Then according to (42), $\mathbb{E}[\log \det(\mathbf{I} + \rho \mathbf{X})] = \mathbb{E}[\log \det(\mathbf{I} + \rho \mathbf{Y})]$.

APPENDIX C

PROOF OF THEOREM 1

We begin by assuming $\mathbf{X} \preceq_c \mathbf{Y}$. By definition, $\mathbb{E}[\log \det(\mathbf{I} + \rho \mathbf{X})] \leq \mathbb{E}[\log \det(\mathbf{I} + \rho \mathbf{Y})]$, $\forall \rho \geq 0$. Using the representation of the log-determinant in terms of the eigenvalues, we get

$$\mathbf{X} \preceq_c \mathbf{Y} \Leftrightarrow \int_0^\infty \log(1 + \rho x) \sum_{i=1}^n f_{\lambda_i(\mathbf{X})}(x) dx \leq \int_0^\infty \log(1 + \rho x) \sum_{i=1}^n f_{\lambda_i(\mathbf{Y})}(x) dx. \quad (43)$$

Now, consider a RV $\lambda_Q(\mathbf{X})$, which is uniformly picked from the set of all eigenvalues of \mathbf{X} . Clearly, its density is given by $n^{-1} \sum_{i=1}^n f_{\lambda_i(\mathbf{X})}(x)$. Therefore, (43) can be rewritten as

$$\mathbf{X} \preceq_c \mathbf{Y} \Leftrightarrow \int_0^\infty \log(1 + \rho x) f_{\lambda_Q(\mathbf{X})}(x) dx \leq \int_0^\infty \log(1 + \rho x) f_{\lambda_Q(\mathbf{Y})}(x) dx, \quad (44)$$

which proves the direct part of the Theorem. The converse can be proved by assuming $\lambda_Q(\mathbf{X}) \leq_c \lambda_Q(\mathbf{Y})$, and retracing the above steps.

REFERENCES

- [1] A. Rajan and C. Tepedelenlioglu, "Ergodic capacity ordering of fading channels," in *Proc. IEEE Int. Symp. Inform. Theory*, 2012.
- [2] A. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Now Publishers Inc, 2004, vol. 1.
- [3] N. Letzepis and A. Grant, "Shannon transform of certain matrix products," in *Proc. IEEE Int. Symp. Inform. Theory*, 2007, pp. 1646–1650.
- [4] J. Quirk and R. Saposnik, "Admissibility and measurable utility functions," *The Review of Economic Studies*, vol. 29, pp. 140–146, 1962.
- [5] F. Belzunce and M. Shaked, "Failure profiles of coherent systems," *Naval Research Logistics*, vol. 51, p. 477, 2004.
- [6] A. Müller and D. Stoyan, *Comparison methods for stochastic models and risks*. John Wiley & Sons Inc, 2002.
- [7] M. Shaked and J. G. Shanthikumar, *Stochastic orders and their applications*, 1st ed. Springer, Oct. 1994.
- [8] C. Tepedelenlioglu, A. Rajan, and Y. Zhang, "Applications of stochastic ordering to wireless communications," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 4249–4257, Dec. 2011.
- [9] R. Schilling, R. Song, and Z. Vondraček, *Bernstein functions: theory and applications*. Walter de Gruyter, 2010.

- [10] N. Lebedev and R. Silverman, *Special functions and their applications*. Dover, 1972.
- [11] R. Bhatia, *Matrix analysis*. Springer Verlag, 1997, vol. 169.
- [12] D. Widder, *The Laplace transform*. Princeton Univ., 1946.
- [13] M. Hasna and M.-S. Alouini, "End-to-end performance of transmission systems with relays over Rayleigh-fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1126 – 1131, Nov. 2003.
- [14] O. Waqar, D. McLernon, and M. Ghogho, "Exact evaluation of ergodic capacity for multihop variable-gain relay networks: A unified framework for generalized fading channels," *IEEE Trans. Vehicular Technol.*, vol. 59, pp. 4181 –4187, Oct. 2010.
- [15] T. Cover and J. Thomas, *Elements of information theory*. Wiley-India, 1999.
- [16] W. Rhee and J. Cioffi, "Ergodic capacity of multi-antenna gaussian multiple-access channels," in *Proc. Thirty-Fifth Asilomar Conf. Signals, Systems and Computers*, vol. 1, 2001, pp. 507–512.

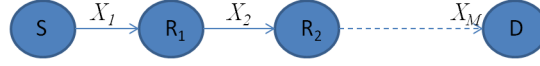


Fig. 1. M -hop relay. S represents the source, R_m represent the relays and D represents the destination.

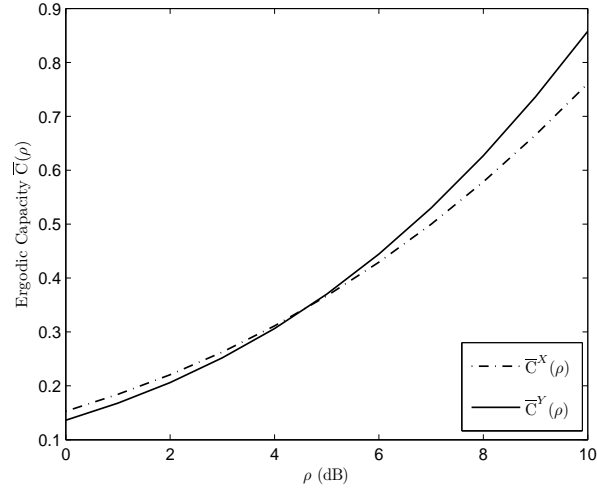


Fig. 2. Ergodic capacity of amplify-forward relay with $M = 3$ slots. The instantaneous SINR is Pareto distributed with parameters $\beta_X = 1$ (dashed line) and $\beta_Y = 3$ (solid line).

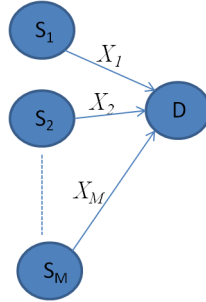


Fig. 3. M -user multiple access channel.